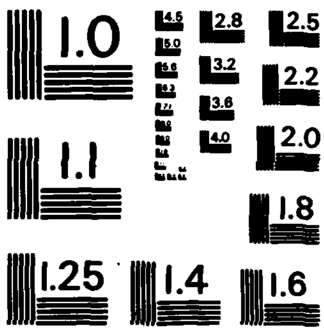
AD-A162 156 ASYMPTOTIC BEHAVIOR OF CONSTRAINED STOCHASTIC APPROXIMATIONS VIA THE THEO..(U) BROWN UNIV PROVIDENCE RI LEFSCHETZ CENTER FOR DYNAMICAL SYSTE.. 1/1

UNCLASSIFIED P DUPUIS ET AL. JUL 85 LCDS-85-12 F/G 12/1 NL

END
FILMED
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6

AFOSR-TR- 85-1053

2

AD-A162 156

Lefschetz Center for Dynamical Systems

Approved for public
distribution unlimi

ILE COPY

2

ASYMPTOTIC BEHAVIOR OF CONSTRAINED STOCHASTIC APPROXIMATIONS VIA THE THEORY OF LARGE DEVIATIONS

by

Paul Dupuis and Harold J. Kushner

July 1985 LCDS Report #85-12

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER AFOSR-TR- 85-1053	2. GOVT ACCESSION NO. AD-A162156
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Asymptotic Behavior of Constrained Stochastic Approximations via the Theory of Large Deviations	5. TYPE OF REPORT & PERIOD COVERED Interim
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Paul Dupuis and Harold J. Kushner	8. CONTRACT OR GRANT NUMBER(s) AFOSR 81-0116
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lefschetz Center for Dynamical Systems Division of Applied Mathematics Brown University, Providence, RI 02912	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A1
11. CONTROLLING OFFICE NAME AND ADDRESS AFOSR Bolling Air Force Base Washington, DC 20332	12. REPORT DATE May 1985
	13. NUMBER OF PAGES 49
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

[redacted]

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ASYMPTOTIC BEHAVIOR OF CONSTRAINED STOCHASTIC APPROXIMATIONS VIA THE THEORY OF LARGE DEVIATIONS

by

Paul Dupuis* and Harold J. Kushner**
Lefschetz Center for Dynamical Systems
Division of Applied Mathematics
Brown University
Providence, Rhode Island 02912

July 1985

* Research has been supported in part by the U.S. Army Research Office under Contract #DAAG 29-84-K-0082, and in part by the Office of Naval Research under Contract #N00014-83-K0542.

** Research has been supported in part by the National Science Foundation Grant #ECS 82-11476, and the Air Force Office of Scientific Research under Contract #[redacted]-AFOSR 81-0116.

Abstract

Let G be a bounded convex set, and Π_G the projection onto G, and $\{\xi_j\}$ a bounded random process. Projected algorithms of the types $X_{n+1}^{\epsilon} = \Pi_G(X_n^{\epsilon} + \epsilon\, b(X_n^{\epsilon}, \xi_n))$ (or $X_{n+1} = \Pi_G(X_n + a_n b(X_n, \xi_n))$, where $0 < a_n \to 0$, $\Sigma a_n = \infty$) occur frequently in applications (among other places) in control and communications theory. The asymptotic convergence properties of $\{X_n^{\epsilon}\}$ as $\epsilon \to 0$, $\epsilon n \to \infty$, have been well analyzed in the literature. Here, we use large deviations methods to get a more thorough understanding of the global behavior. Let θ be a stable point of the algorithm in the sense that $X_n^{\epsilon} \to \theta$ in distribution as $\epsilon \to 0$, $n\epsilon \to \infty$. For the unconstrained case, rate of convergence results involve showing asymptotic normality of $\{(X_n^{\epsilon} - \theta)/\sqrt{\epsilon}\}$, and use linearizations about θ. In the constrained case θ is often on ∂G, and such methods are inapplicable. But the large deviations method yields an alternative which is often more useful in the applications. The action functionals are derived and their properties (lower semicontinuity, etc.) are obtained. The statistics (mean value, etc.) of the escape times from a neighborhood of θ are obtained, and the global behavior on the infinite interval is described.

I. Introduction

Let $q_i(\cdot)$, $i \leq k$, be continuously differentiable and let $G = \{x: q_i(x) \leq 0, i \leq k\}$ be a compact convex set which is the closure of its interior. Let $\{\xi_n\}$ be a bounded sequence of random variables and $b(\cdot,\cdot)$ a bounded function with $b(\cdot,\xi)$ uniformly (in ξ) Lipschitz continuous. Define $\Pi_G(x)$ to be the (nearest point) projection of x onto G.

This document describes a ... which

The projected recursive (or stochastic approximation) algorithm ~~(1.1)~~ arises frequently in applications in control and communications theory.

$$X^\epsilon_{n+1} = \Pi_G(X^\epsilon_n + \epsilon b(X^\epsilon_n, \xi_n)), \quad x \in R^r, X^\epsilon_0 = x_0 \text{ given} \tag{1.1}$$

There is a sizeable literature (e.g., [1] to [5]) concerning its asymptotic properties as $\epsilon \to 0$ with $\epsilon n \to t$ or $\epsilon n \to \infty$. Often ϵ is replaced by a 'stochastic approximation' sequence $\{a_n\}$ with $a_n \to 0$, $a_n > 0$, and $\Sigma a_n = \infty$. The methods of analysis are similiar in both cases, except that the latter case $(a_n \to 0)$ allows the possibility of w.p.1 convergence of $\{X_n\}$.

Typical results are the following. For a velocity vector v, define the projection of v at $x \in G$ by $\Pi_G(x,v) = \lim_{\Delta} [\Pi_G(x + \Delta v) - x]/\Delta$ and write $\bar{b}(x) = Eb(x,\xi)$ ($\bar{b}(\cdot)$ to be redefined below). The equation

$$\dot{x} = \Pi_G(x, \bar{b}(x)) \tag{1.2}$$

represents the projected dynamics on G for the ODE $\dot{x} = \bar{b}(x)$. Let $x^\epsilon(\cdot)$

Epsilon approaching limit of 0 with epsilon sub n approaching limit of t or epsilon sub n approaching infinity. Keywords: Convergence; asymptotic normality.

denote the piecewise linear interpolation of $\{X_n^\epsilon\}$ with interpolation interval ϵ. Under reasonable conditions, X_n^ϵ converges in distribution to the set of stationary points of (1.2), as $\epsilon \to 0$ and $\epsilon n \to \infty$; also $x^\epsilon(\cdot) \to x(\cdot)$, a process which satisfies (1.2). If $\bar{b}(x)$ is a gradient of the function $-\bar{B}(x)$, then the limit points are the Kuhn-Tucker points for the problem of minimizing $\bar{B}(\cdot)$ on G. Rate of convergence results for (1.1) are unavailable. For the unconstrained case the 'rate' results are of the following form. Let $X_n^\epsilon \to \theta$ in distribution as $\epsilon \to 0$, $\epsilon n \to \infty$. Define $U_n^\epsilon = (X_n^\epsilon - \theta)/\sqrt{\epsilon}$, and let $U^\epsilon(\cdot)$ denote the continuous parameter interpolation (interval ϵ). Then, under the appropriate conditions, $U^\epsilon(t_\epsilon+\cdot)$ converges weakly to a stationary Gauss-Markov process as $\epsilon \to 0$ if $t_\epsilon \to \infty$ fast enough [12] (with a similar result for the stochastic approximation case). The result is based on a local linearization about θ, and the rate result does not fully exploit the dynamics of the iteration. Such a linearization cannot, in any case, be done for (1.1) when the limit θ is on the boundary ∂G. Some results for this case are in [5], where $\{X_n^\epsilon\}$ is Markov, and (under appropriate conditions) $(X^\epsilon - \theta)/\epsilon$ is shown to converge in distribution as $\epsilon \to 0$, where X^ϵ is distributed according to a (unique) invariant measure for (1.1).

Here, we use the theory of large deviations to get a better picture of the asymptotic properties for (1.1). Let D denote a neighborhood of θ (all neighborhoods are with respect to G), with θ a stable point of (1.2) and with $\bar{D}$ in the domain of attraction of θ. Let $\tau_D^\epsilon = \min\{t: x^\epsilon(t) \notin D\}$, define $C_x[0,T]$ to be the set of G-valued continuous functions on $[0,T]$ with initial condition x, and let P_x denote the probability measure given that $x_0 = x$. We always use $d(\cdot,\cdot)$ to denote the (sup norm) distance between functions in $C_x[0,T]$, as well as the sup norm distance between points in a Euclidean space . As special cases of our large deviations results, we obtain estimates for quantities such as

$$\lim_{\epsilon} \epsilon \log P_x\{\tau_D^\epsilon \leq T\} \tag{1.3}$$

$$\lim_{\epsilon} \epsilon \log P_x\{x^\epsilon(\cdot) \in A\} , A \subset C_x[0,T] , \tag{1.4}$$

$$\lim_{\epsilon} \epsilon \log E_x \tau_D^\epsilon .$$

The limits in (1.3, 1.4) are important in studying the asymptotic properties of $\{X_n^\epsilon\}$ and are often of greater interest than 'local' results of the type of limits of suitably normalized $(X_n^\epsilon - \theta)$. 'We can obtain the (asymptotic) locations of the exit from D and the most likely escape routes, all of which are important in applications. A comparison of (1.3) for different algorithms yields information on their relative stability. They exploit more of the structure of the algorithm than the 'local' limits do, and often provide realistic information, e.g., estimates of the time spent in a neighborhood of a stable point, etc.

The paper is organized as follows. In Section 2, various terms from the theory of large deviations are introduced, and the problem (on a time interval [0,T]) formulated. Sections 3 and 4 contain some technical results concerning the action functional and approximations of (1.1). These are put together in Section 5 to get the general large deviations result. Section 6 concerns the mean escape time of (1.1) from a neighborhood of a stable point of (1.2) and in Section 7 we remark on some extensions to the global behavior of (1.1) on the infinite time interval $[0,\infty]$, the character of movement from stable point to stable point and on the stochastic approximation case.

II. Problem Formulation, Assumptions and Definitions

Let T/Δ be an integer and define $I_i = \{j: i\Delta/\epsilon \leq j < (i\Delta+\Delta)/\epsilon\}$. Suppose that the limit (defining $\bar{b}(\cdot)$) in (2.1) exists uniformly for $x \in G$:

$$\text{(2.1)} \qquad \lim_N \frac{1}{N} \sum_0^{N-1} E\, b(x,\xi_i) = \bar{b}(x) .$$

Suppose that there is a function $H(\cdot,\cdot)$ such that for each $\Delta > 0$, the limit in (2.2) exists uniformly for (x_i,α_i) in any compact set.

$$\text{(2.2)} \qquad \sum_{i=0}^{T/\Delta-1} \Delta\, H(\alpha_i,x_i) = \lim_N \frac{\Delta}{N} \log E \exp \sum_{i=0}^{T/\Delta-1} \alpha_i' \sum_{j=iN}^{iN+N-1} b(x_i,\xi_j).$$

$H(\cdot,\cdot)$ is obviously continuous, and we suppose that $H(\cdot,x)$ is continuously differentiable. The limits exist and we have the differentiability in α if $\{\xi_i\}$ is a finite state ergodic Markov chain (see [7], where the argument is based on one in [6]) or if $\xi_n = \sum_k g_{n-k}\psi_k$ and $\{\psi_k\}$ is i.i.d., bounded, $\Sigma k|g_k| < \infty$ and $g_k = 0$ for $k < 0$ [8]. Define the usual Legendre transform L and action functional S by

$$L(\beta,x) = \sup_\alpha[\beta'\alpha - H(\alpha,x)]$$

$$S(T,\phi) = \int_0^T L(\dot{\phi}(s), \phi(s))ds \quad \text{for } \phi(\cdot) \text{ absolutely continuous } \in C_x[0,T],$$

$$= \infty \quad \text{otherwise}$$

$L(\cdot,\cdot)$ is lower semicontinuous (l.s.c.), $L(\cdot,x)$ is convex, and $S(T,\cdot)$ is l.s.c. [6]. The sets $U(x) = \{\beta: L(\beta,x) < \infty\}$ and $U_0(x) = U(x) - \bar{b}(x)$ are convex and are uniformly bounded since $b(x,\xi)$ is uniformly bounded [7]. Assume that $U(\cdot)$ is continuous in the Hausdorff topology.

Define $B(x,\beta) = \{v: \Pi_G(x,v) = \Pi_G(x,\beta)\}$, the set of 'velocities' having the same projection at x as β has. $B(\cdot,\cdot)$ is upper semicontinuous (u.s.c.) in the Hausdorff topology. Define

(2.3) $$L_G(\beta,x) = \inf_{v\in B(x,\beta)} L(v,x) .$$

By the l.s.c. of $L(\cdot,\cdot)$ and u.s.c. of $B(\cdot,\cdot)$, $L_G(\cdot,\cdot)$ is l.s.c. For $\phi(\cdot)$ absolutely continuous, set

(2.4) $$S_G(T,\phi) = \int_0^T L_G(\dot{\phi}(s),\phi(s))ds$$

and define $S_G(T,\phi) = \infty$ otherwise.

Under some other conditions to be introduced below, the main result of the next three sections is that $S_G(T,\phi)$ is an action functional for $x^\epsilon(\cdot)$ in that $S_G(T,\cdot)$ is l.s.c., and for $A \subset C_x[0,T]$ (with interior A° and closure $\bar{A}$),

(2.5) $$-\inf_{\phi\in A^\circ} S_G(T,\phi) \le \underline{\lim}_{\epsilon}\, \epsilon \log P_x\{x^\epsilon(\cdot) \in A\}$$

$$\le \overline{\lim}_{\epsilon}\, \epsilon \log P_x\{x^\epsilon(\cdot) \in A\} \le -\inf_{\phi\in \bar{A}} S_G(T,\phi)$$

Owing to the presence of the boundary, the analysis is somewhat non-standard. The typical results of interest require that the boundary be taken into account. See, e.g., Fig. 1. where D is a set in the domain of attraction of θ, a stable point of (1.2), and the arrows indicate the flow lines for (1.2). With the indicated 'typical' $U(x)$, the most likely escape paths from D are along the boundary $[a,b]$.

To relate (2.5) to (1.3), let $A = \{\phi(\cdot)\colon \phi(0) = x,\ \phi(t) \notin D$ for some $t \leq T\}$.

Write the vectors x, b, β, etc. as (x_1,x_2), (b_1,b_2), (β_1,β_2), etc., where x_1, b_1, have dimension r_1 and x_2, b_2, have dimension r_2 for some r_1, r_2. Σ_{11} (resp., Σ_{22}) below is $r_1 \times r_1$ (resp., $r_2 \times r_2$). For the purposes of simplifying the analysis from Section 4 on, we make an additional assumption. Define $\Sigma(\cdot)$ and $\Sigma_N(\cdot)$ by

$$\lim \frac{1}{N} \operatorname{cov} \sum_1^N [b(x,\xi_j) - \bar{b}(x)] \equiv \lim \Sigma_N(x) = \Sigma(x) \tag{2.6}$$

$$= \begin{bmatrix} \Sigma_{11}(x) & \Sigma_{12}(x) \\ \Sigma_{21}(x) & \Sigma_{22}(x) \end{bmatrix}$$

We use either of the two cases. <u>Case</u> 1 (<u>non-degenerate</u>), where $\Sigma(x)$ is positive definite on G. <u>Case</u> 2 (<u>degenerate</u>), where $\Sigma_{11}(x) = \Sigma_{21}(x) = \Sigma_{12}(x) = 0$, and $\Sigma_{22}(x)$ is positive definite on G. These cover the typical cases in applications. In Case 2, $L(\beta,x) = \infty$ unless $\beta_1 = \bar{b}_1(x)$. Define $U_2(x) = \{\beta_2\colon$

$L(\bar{b}_1(x),\beta_2,x) < \infty\}$ and define the δ-interior sets $U^\delta(x) = \{\beta \in U(x): d(\beta,\partial U(x)) \geq \delta\}$, $U_2^\delta(x) = \{\beta_2 \in U_2(x): d(\beta_2,\partial U_2(x)) \geq \delta\}$.

Non-degenerate case. $H(\cdot,x)$ is strictly convex in a neighborhood of $\alpha = 0$, uniformly in x in G, and for any $\delta > 0$ $L(\cdot,\cdot)$ is uniformly continuous for $(\beta,x) \in \{U^\delta(x),\ x \in G\}$. Also $L(\beta,x) = 0$ iff $\beta = \bar{b}(x)$, and there is a neighborhood N of the origin such that $N + \bar{b}(x) \subset U(x)$ for all $x \in G$ [7], and $L(\bar{b}(x) + \cdot,x)$ is strictly convex on N, uniformly in x in G.

Degenerate Case. Here the definition (2.2) reduces to

$$H(\alpha,x) = \alpha_1' \bar{b}_1(x) + H_2(\alpha_2,x),$$

where $H_2(\cdot,\cdot)$ is defined by (write $\alpha_i = (\alpha_{1i},\alpha_{2i})$)

$$\sum_{i=0}^{N/\Delta-1} \Delta\, H_2(\alpha_{2i},x_i) = \lim_N \frac{\Delta}{N} \log E \exp \sum_{i=0}^{T/\Delta-1} \alpha_{2i}' \sum_{j\in I_i} b_2(x_i,\xi_j) \quad . \tag{2.7}$$

Let $L_2(\beta_2,x)$ be the dual of $H_2(\alpha_2,x)$. Then $H_2(\cdot,\cdot)$ and $L_2(\cdot,\cdot)$ have the properties ascribed above to $H(\cdot,\cdot)$ and $L(\cdot,\cdot)$; also $L(\beta,x) = L_2(\beta_2,x)$ if $\beta_1 = \bar{b}_1(x)$ and $L(\beta,x) = \infty$ otherwise. The following result will be useful.

<u>Lemma 1.</u> <u>Let</u> $v^\delta \in U^\delta(x)$ <u>and</u> $v^\delta \to v$. <u>Then</u> $L(v^\delta,x) \to L(v,x)$. <u>For</u> $\delta > 0$,

$$L(\bar{b}(x) + (1-\delta)v,x) \leq L(\bar{b}(x) + v,x), \quad x \in G, \text{ all } v. \tag{2.8}$$

The first assertion is in Freidlin [6]. The second is a consequence of the convexity of $L(\cdot,x)$ and the fact that $L(\bar{b}(x),x) = 0$, $L(b(x) + v,x) \geq 0$.

III. Discrete Approximations: Preliminary Formulation

Owing to the boundary ∂G, it is hard to get the large deviations results for (1.1) directly. We do it in a sequence of approximations, which get 'closer' to (1.1), and for each of which we can get a large deviations result from the preceeding one. We define the approximations in this section. $\psi(\cdot)$ and $\phi(\cdot)$ denote arbitrary functions in $C_x[0,T]$. For $\Delta > 0$ (w.l.o.g. we let T/Δ and Δ/ϵ be integers, with $N\Delta = T$) define ψ_j^Δ to equal $\psi(n\Delta)$ for $j \in I_n$, and define the sequence $\{Y_i^{\epsilon,\psi,\Delta}, i \leq N\}$ by $Y_0^{\epsilon,\psi,\Delta} = x$ and

$$Y_{i+1}^{\epsilon,\psi,\Delta} = Y_i^{\epsilon,\psi,\Delta} + \epsilon \sum_{j=I_i} b(\psi_j^\Delta, \xi_j) \tag{3.1}$$

Let $\psi^\Delta(\cdot)$ denote the piecewise constant (on intervals $[n\Delta, n\Delta + \Delta)$) interpolation of $\{\psi(i\Delta)\}$. We use ψ^Δ to represent the samples $\{\psi(i\Delta), 0 < i \leq N\}$. We first state a large deviations result for (3.1), then use the 'contraction' principle to get such a result for a projected form of (3.1), and then take appropriate limits as $\Delta \to 0$. In [6], Freidlin developed the large deviations theory for (3.1). The details in [6] were for a continuous parameter case, but the results and methods would be identical for the discrete parameter case (3.1). In particular, the following results hold. Define $S^{\psi,\Delta}$ by

$$S^{\psi,\Delta}(T,\phi) = \sum_0^{N-1} \Delta L\left(\frac{\phi(i\Delta+\Delta) - \phi(i\Delta)}{\Delta}, \psi(i\Delta)\right). \tag{3.2}$$

Then $S^{\psi,\Delta}(T,\phi)$ is an action functional for $\{Y_i^{\epsilon,\psi,\Delta}, i \leq N\}$ in the sense that for any Borel set B in $(R^r)^N$

$$(3.3)\qquad -\inf_{\phi^\Delta \in B^\circ} S^{\psi,\Delta}(T,\phi) \leq \underline{\lim}_{\epsilon}\ \epsilon \log P_x\left\{(Y_i^{\epsilon,\psi,\Delta}, 0 < i \leq N) \in B\right\}$$

$$\leq \overline{\lim_{\epsilon}}\ \epsilon \log P_x\left\{(Y_i^{\epsilon,\psi,\Delta}, 0 < i \leq N) \in B\right\}$$

$$\leq -\inf_{\phi^\Delta \in \bar{B}} S^{\psi,\Delta}(T,\phi) .$$

For $\phi(\cdot) \in C_x[0,T]$, define

$$(3.4)\qquad S_G^{\psi,\Delta}(T,\phi) = \inf_f S^{\psi,\Delta}(T,f),$$

<u>where the</u> inf <u>is over the set</u>

$$\{f: \Pi_G(\phi(i\Delta) + f(i\Delta+\Delta) - f(i\Delta)) = \phi(i\Delta+\Delta), i \leq N-1\}.$$

For later use, it is more convenient to rewrite (3.4) in the form

$$(3.5)\qquad S_G^{\psi,\Delta}(T,\phi) = \sum_0^{N-1} \Delta \inf_f L\left[\frac{f(i\Delta+\Delta) - f(i\Delta)}{\Delta}, \psi(i\Delta)\right],$$

where the inf is over the same set as for (3.4). By the 'contraction principle' ([9],p5), $\overset{f}{S}{}_G^{\psi,\Delta}(T,\phi)$ is the action functional for the 'projected' sequence $\{X_i^{\epsilon,\psi,\Delta},\ i \leq N\}$ defined by $X_0^{\epsilon,\psi,\Delta} = x$ and

$$X_{i+1}^{\epsilon,\psi,\Delta} = \Pi_G\,(X_i^{\epsilon,\psi,\Delta} + \epsilon \sum_{j\in I_i} b(\psi_j^{\Delta}, \xi_j))\ . \tag{3.6}$$

In the next section we prove (Theorem 1) that $S_G^{\phi,\Delta}(T,\phi) \equiv S_G^{\Delta}(T,\phi)$ is an action functional for the next approximation $\{X_i^{\epsilon,\Delta},\ i \leq N\}$ defined by $X_o^{\epsilon,\Delta} = x$ and

$$X_{i+1}^{\epsilon,\Delta} = \Pi_G\,(X_i^{\epsilon,\Delta} + \epsilon \sum_{j\in I_i} b(X_i^{\epsilon,\Delta}, \xi_j))\ . \tag{3.7}$$

Let $x^{\epsilon,\Delta}(\cdot)$ denote the piecewise constant interpolation of $\{X_i^{\epsilon,\Delta}\}$ (interpolation interval Δ). For a set $A \subset C_x[0,T]$ it will sometimes be convenient to use the 'sampled' notation $x^{\epsilon,\Delta}(\cdot) \in A^{\Delta}$ to mean that $X_i^{\epsilon,\Delta} = \phi(i\Delta)$ for $i \leq N$ for some $\phi(\cdot) \in A$.

In Theorem 2, the l.s.c. of $S_G(T,\phi)$ is proved, as is the relation

$$\begin{aligned} -\inf_{\phi\in A^o} S_G(T,\phi) &\leq \underline{\lim}_{\Delta}\ \underline{\lim}_{\epsilon}\ \epsilon \log P_x\{x^{\epsilon,\Delta}(\cdot) \in A^{\Delta}\} \\ &\leq \overline{\lim}_{\Delta}\ \overline{\lim}_{\epsilon}\ \epsilon \log P_x\{x^{\epsilon,\Delta}(\cdot) \in A^{\Delta}\} \\ &\leq -\inf_{\phi\in \bar{A}} S_G(T,\phi)\ . \end{aligned} \tag{3.8}$$

In Theorem 3 we show that $x^{\epsilon}(\cdot)$ and $x^{\epsilon,\Delta}(\cdot)$ are 'close' in the following sense. Let $d(x^{\epsilon}(\cdot),\phi(\cdot)) \leq \delta$. Then there is a $\delta_1(\delta_1 \to 0$ as $\delta \to 0)$ such that $d(x^{\epsilon,\Delta}(\cdot),\phi(\cdot)) \leq \delta_1$, and conversely. This result and (3.8) will enable us to get the desired large deviations result in Theorem 4.

IV. Properties of $S_G^{\Delta}(T,\phi)$.

Write $X^{\epsilon,\Delta} = \{X_i^{\epsilon,\Delta},\ 0 < i \leq N\}$, $X^{\epsilon,\psi,\Delta} = \{X_i^{\epsilon,\psi,\Delta},\ 0 < i \leq N\}$, $\psi^{\Delta} = \{\psi(i\Delta),\ 0 < i \leq N\}$.

Theorem 1. *For each* $\Delta > 0$, $S_G^{\Delta}(T,\phi)$ *is an action functional for* $X^{\epsilon,\Delta}$; *i.e., for any Borel set* B *in* $(R^r)^N$,

$$
\begin{aligned}
-\inf_{\phi^{\Delta}\in B^0} S_G^{\Delta}(T,\phi) \leq \varliminf_{\epsilon} \epsilon \log P_X\{X^{\epsilon,\Delta}\in B\} &\leq \varlimsup_{\epsilon} \epsilon \log P_X\{X^{\epsilon,\Delta}\in B\} \\
&\leq -\inf_{\phi^{\Delta}\in\bar{B}} S_G^{\Delta}(T,\phi)\,. \qquad (4.1)
\end{aligned}
$$

Proof. Given $\delta > 0$, there is a $\delta_1 > 0$ (where $\delta_1 \to 0$ as $\delta \to 0$) such that

$$d(X^{\epsilon,\psi,\Delta},\psi^{\Delta}) < \delta_1 \Rightarrow d(X^{\epsilon,\Delta},\psi^{\Delta}) < \delta\,.$$

To see this, write $\psi(i\Delta) = X_i^{\epsilon,\psi,\Delta} + \alpha_i$, where $|\alpha_i| \leq \delta_1$ and

$$X_{i+1}^{\epsilon,\psi,\Delta} = \Pi_G\Big(X_i^{\epsilon,\psi,\Delta} + \epsilon \sum_{j\in I_i} b(X_i^{\epsilon,\psi,\Delta} + \alpha_i, \xi_j)\Big) \qquad (4.2a)$$

$$X_{i+1}^{\epsilon,\Delta} = \Pi_G\Big(X_i^{\epsilon,\Delta} + \epsilon \sum_{j\in I_i} b(X_i^{\epsilon,\Delta},\xi_j)\Big)\,. \qquad (4.2b)$$

The existence of a suitable δ_1 follows from this and the relation

$$|\Pi_G(x) - \Pi_G(x')| \leq |x-x'| \quad \text{(due to convexity of } G)$$

and the uniform Lipschitz condition on $b(\cdot,\xi)$.

Similarly there is a $\delta_2 > 0$ (where $\delta_2 \to 0$ as $\delta \to 0$) such that

$$d(X^{\epsilon,\Delta}, \psi^\Delta) < \delta \Rightarrow d(X^{\epsilon,\psi,\Delta}, \psi^\Delta) < \delta_2 .$$

Thus

$$P_x\{d(X^{\epsilon,\Delta}, \psi^\Delta) < \delta\} \geq P_x\{d(X^{\epsilon,\psi,\Delta}, \psi^\Delta) < \delta_1\}, \tag{4.3}$$

$$P_x\{d(X^{\epsilon,\psi,\Delta}, \psi^\Delta) < \delta_2\} \geq P_x\{d(X^{\epsilon,\Delta}, \psi^\Delta) < \delta\} . \tag{4.4}$$

The result follows from (4.3) - (4.4) in the standard way [6]. In particular, given B (with non-empty B^0) and $h > 0$, there are $\psi(\cdot)$ with $\psi^\Delta \in B^0$ and small δ, δ_1, such that

$$\begin{aligned} P_x\{X^{\epsilon,\Delta} \in B\} &\geq P_x\{X^{\epsilon,\Delta} \in B^0\} \geq \\ &P_x\{d(X^{\epsilon,\Delta}, \psi^\Delta) < \delta\} \geq \\ &P_x\{d(X^{\epsilon,\Delta,\psi}, \psi^\Delta) < \delta_1\} \geq \\ &\exp -\tfrac{1}{\epsilon}[S_G^{\Delta,\psi}(T,\psi) + h] \end{aligned}$$

for small ϵ. The left side of (4.1) follows from an appropriate choice of ψ (a 'tail' element of the infimizing sequence). The right hand inequality of (4.1) follows from a similar approximation arguement. Q.E.D.

Theorem 2. $S_G(T,\cdot)$ is l.s.c. For each $A \in C_x[0,T]$,

$$\underline{\lim_{\Delta}} \inf_{\phi \in \bar{A}^{\Delta}} S_G^{\Delta}(T,\phi) \geq \inf_{\phi \in \bar{A}} S_G(T,\phi). \tag{4.5}$$

For each $\phi(\cdot)$ for which $S_G(T,\phi) < \infty$, there are piecewise constant (on intervals of length Δ) functions $\phi_\Delta(\cdot)$, $\psi_\Delta(\cdot)$ converging uniformly to $\phi(\cdot)$ such that

$$\overline{\lim_{\Delta}} \; S_G^{\psi_\Delta,\Delta}(T,\phi_\Delta) \leq S_G(T,\phi). \tag{4.6}$$

The inequalities (3.8) hold.

Proof. Part I. The proofs for the degenerate and non-degenerate cases are essentially the same and we do the latter case only (for the degenerate case, use U_2^δ in lieu of U^δ below).

Proof of the l.s.c. of $S_G(T,\cdot)$. Let $\phi_n(\cdot) \to \phi(\cdot)$. The infimizing v is attained in $\inf_{v \in B(\phi^n(s),\dot{\phi}^n(s))} L(v,\dot{\phi}^n(s))$. Let $\bar{v}^n(\cdot)$ be a measurable selection

*[11, Thm.4.1] of the minimizer and write it as $\tilde{v}^n(s) = v^n(s) + \bar{b}(\phi^n(s))$. Using the uniform continuity of $L(\cdot,\cdot)$ on $\{\beta, x : \beta \in U^\delta(x),\ x \in G\}$ for $\delta > 0$ we have

$$\underline{\lim}_{n}\ S_G(T,\phi_n) = \underline{\lim}_{n} \int_0^T L\left[\bar{b}(\phi^n(s)) + v^n(s), \phi^n(s)\right]ds$$

$$(4.7) \qquad \geq \underline{\lim}_{\delta}\ \underline{\lim}_{n} \int_0^T L\left[\bar{b}(\phi^n(s)) + (1-\delta)v^n(s), \phi^n(s)\right)ds$$

$$= \underline{\lim}_{\delta}\ \underline{\lim}_{n} \int_0^T L\left[\bar{b}(\phi(s)) + (1-\delta)v^n(s), \phi(s)\right]ds$$

$$= \underline{\lim}_{\delta}\ \underline{\lim}_{\Delta} \left\{ \underline{\lim}_{n} \int_0^T L\left[\bar{b}(\phi(s)) + (1-\delta)v^n(s), \phi^\Delta(s)\right]ds \ - \alpha_\delta^\Delta \right\}$$

$$\geq \underline{\lim}_{\delta}\ \underline{\lim}_{\Delta} \left\{ \underline{\lim}_{n} \sum_0^{N-1} \Delta\, L\left(\frac{1}{\Delta}\int_{i\Delta}^{i\Delta+\Delta} [\,\bar{b}(\phi(s)) + (1-\delta)\, v^n(s)\,]ds, \phi(i\Delta)\right) - \alpha_\delta^\Delta \right\},$$

where $\alpha_\delta^\Delta \to 0$ as $\Delta \to 0$ for each $\delta > 0$. The first inequality uses Lemma 1, and the last inequality follows from Jensens' inequality and the convexity of $L(\cdot,x)$.

---

* The selection theorem 4.1 in [11] uses a bounded u.s.c. function and a maximization. But a slight modification works for our case, since the l.s.c. $L(\cdot,\cdot)$ is bounded from below.

Choose a subsequence such that $\int_0^t v^n(s)ds$ converges, with limit denoted by (absolutely continuous since the $U(x)$ are bounded) $V(\cdot)$, and write $V(t) = \int_0^t v(s)ds$. By the l.s.c. of $L(\cdot,\cdot)$ and Fatou's lemma, we can continue the string of inequalities in (4.7) as

$$(4.7') \qquad \geq \frac{\lim}{\delta} \int_0^T L(\bar{b}(\phi(s)) + (1-\delta)v(s),\phi(s))ds$$

$$\geq \int_0^T L(\bar{b}(\phi(s)) + v(s),\phi(s))ds \quad .$$

If $\Pi_G(\phi(s),\bar{b}(\phi(s)) + v(s)) = \dot{\phi}(s)$ for almost all $s \leq T$ we are done, since in that case (for almost all s) $\bar{b}(\phi(s) + v(s)) \in B(\phi(s),\dot{\phi}(s))$ and

$$(4.8) \qquad \begin{aligned} &L(\bar{b}(\phi(s)) + v(s),\phi(s)) \\ &\quad \geq \inf_{v \in B(\phi(s),\dot{\phi}(s))} L(v,\phi(s)) = L_G(\phi(s), \dot{\phi}(s)) \ . \end{aligned}$$

Thus, we need only show the projection property below (4.7') for $b(\phi(s)) + v(s) = \bar{v}(s)$.

If for some $s < T$, $\dot{\phi}(s) \in G^0$, the interior of G, then $\bar{v}^n(t) = \dot{\phi}^n(t)$ for almost all t and large n, on some open interval containing s. This implies that $\bar{v}(s) = \dot{\phi}(s)$ for almost all s such that $\phi(s) \in G^0$. Now, let $\phi(s) \in \partial G$ on some interval. In particular let $I = [a,b]$, $a < b$, be such that (rearrange the indices is necessary) for some $\delta > 0$, and integer ℓ and all $s \in I$, $q_i(\phi(s)) = 0$, $i \leq \ell$, $q_i(\phi(s)) \leq -\delta < 0$, $i > \ell$. Define the set $G(\phi) = \{y:$

$q_i(y) \leq 0,\ i \leq \ell\}$. Let $C(x)$ denote the cone generated by the outer normals to $\{y: q_i(y) \leq 0\}$, $i \leq \ell$, at the point x. Then $C(\phi^n(s)) \to C(\phi(s))$ for $s \in I$.

Define the 'projection error' $\hat{v}^n(s)$ by

$$\bar{v}^n(s) - \hat{v}^n(s) = \Pi_{G(\phi)}(\phi^n(s), \bar{v}^n(s)) = \dot{\phi}^n(s).$$

Then $\hat{v}^n(s) \in C(\phi^n(s))$, if $\phi^n(s) \in \partial G(\phi)$. Otherwise $\hat{v}^n(s) = 0$. Extracting a convergent subsequence if necessary, there is an absolutely continuous function $\hat{V}(\cdot)$ such that

$$\int_a^t \hat{v}^n(s)ds \to \hat{V}(t) \equiv \int_a^t \hat{v}(s)ds,\ t \leq b\ .$$

Note that $\phi(\cdot)$ moves orthogonally to $C(\phi(s))$ at s (recall that the active constraints for $\phi(s)$ do not change for $s \in I$). Since $\phi^n(\cdot) \to \phi(\cdot)$ and $\phi(s) \in \partial G(\phi)$ on I,

$$\phi^n(a) + \int_a^t [\bar{v}^n(s) - \hat{v}^n(s)]ds \to \phi(a) + \int_a^t [\bar{v}(s) - \hat{v}(s)]ds = \phi(t),\ t \leq b.$$

Thus $\bar{v}(s) - \hat{v}(s) = \dot{\phi}(s)$ for a.a. $s \in I$. By construction, $\hat{v}(s) \perp \bar{v}(s) - \hat{v}(s)$ and $\hat{v}(s) \in C(\phi(s))$ for almost all $s \in I$. Thus $\Pi_G(\phi(s), \bar{v}(s)) = \Pi_G(\phi(s), \bar{v}(s) - \hat{v}(s) + \hat{v}(s)) = \Pi_G(\phi(s), \bar{v}(s) - \hat{v}(s)) = \Pi_G(\phi(s), \dot{\phi}(s))$. a.a. $s \in I$. By this method, we can show that $\dot{\phi}(s) = \Pi_G(\phi(s), \bar{v}(s))$ for a.a. $s \in [0,T]$ and the 'projection' requirement below (4.7$'$) holds. Thus $S_G(T,\cdot)$ is l.s.c.

Part 2. We can write

$$S_G^{\Delta}(T,\phi) = \int_0^T \inf_u \; L(u, \phi^{\Delta}(s))ds \quad ,$$

where on the interval $[i\Delta, i\Delta + \Delta)$, $\inf_u$ is the $\inf$ over all u such that

$$\frac{\Pi_G(\phi(i\Delta) + \Delta u) - \phi(i\Delta)}{\Delta} = \frac{\phi(i\Delta + \Delta) - \phi(i\Delta)}{\Delta}$$

By using this and an argument very similar to that used to get the l.s.c. property in Part 1, we can show that

$$\text{(4.9)} \qquad \underline{\lim}_{\Delta} \, S_G^{\Delta}(T,\phi) \geq S_G(T,\phi).$$

Also, if $\phi_\Delta(\cdot) \to \phi(\cdot)$ ($\phi_\Delta(\cdot)$ being piecewise constant), a similar argument yields

$$\text{(4.10)} \qquad \underline{\lim}_{\Delta} \, S_G^{\Delta}(T,\phi_\Delta) \geq S_G(T,\phi).$$

We now prove (4.5). Let $\inf_{\phi \in \bar{A}} S_G(T,\phi) < \infty$ and let $\phi_0(\cdot)$ attain the inf. Let (piecewise constant) $\phi_\Delta(\cdot)$ yield the inf in $\inf_{\phi \in A^{\Delta}} S_G^{\Delta}(T,\phi)$. Since $\{(\phi_\Delta(i\Delta + \Delta) - \phi_\Delta(i\Delta))/\Delta, \; i\Delta \leq T\}$ is bounded, we can extract a convergent subsequence of the piecewise constant functions $\{\phi_\Delta(\cdot)\}$ with an absolutely continuous limit $\phi(\cdot)$. Then by (4.10)

$$\underline{\lim_{\Delta}}\ S_G^{\Delta}(T,\Phi_\Delta) \geq S_G(T,\Phi) \geq S_G(T,\Phi_0)\ ,$$

which implies that (4.5) holds, together with the right side of (3.8). If $\inf_{\Phi\in\bar{A}} S_G(T,\Phi) = \infty$, then the above argument yields that $\underline{\lim}_{\Delta}\ S_G^{\Delta}(T,\Phi_\Delta) = \infty$ also.

Inequality (4.6) yields the left side (3.8), by use of the following observations. (See Part 1 for a related argument.) If A^0 is not empty, then there is a 'nearly infimizing' $\Phi(\cdot) \in A^0$ and a $\delta > 0$ such that for small ϵ,Δ,

$$\begin{aligned} &P_x\{(x^{\epsilon,\Delta}(\cdot) \in A^{\Delta}\} \geq P_x\{x^{\epsilon,\Delta}(\cdot) \in (A^{\Delta})^0\} \\ (4.11a) \qquad &\geq P_x\{d(x^{\epsilon,\Delta}(\cdot),\Phi(\cdot)) < \delta\}\ . \end{aligned}$$

Given $h > 0$ there is a $\delta_1 > 0$ and $\epsilon_0 > 0$ such that for $\epsilon \leq \epsilon_0$ and small Δ and $\tilde{\Psi}_\Delta(\cdot),\tilde{\Phi}_\Delta(\cdot)$ close enough to $\Phi(\cdot)$, we can continue the inequalities as (using (4.6))

$$\begin{aligned} &\geq P\{d(x^{\epsilon,\tilde{\Psi}_\Delta,\Delta}(\cdot),\ \tilde{\Phi}_\Delta(\cdot)) < \delta_1\} \\ (4.11b) \qquad &\geq \exp -\tfrac{1}{\epsilon}\left[S_G^{\tilde{\Psi}_\Delta,\Delta}(T,\tilde{\Phi}_\Delta) + h\right] \\ &\geq \exp -\tfrac{1}{\epsilon}\left[S_G(T,\Phi) + 2h\right]. \end{aligned}$$

Thus, to get the l.h.s. of (3.8) only (4.6) needs to be proved. To do this, we adapt an argument of Freidlin [6].

Part 3. Proof of (4.6). We can write

$$S_G(T,\Phi) = \int_0^T \inf_{v \in B(\Phi(s),\dot{\Phi}(s))} \sup_{\alpha} [\alpha' v - H(\alpha,\Phi(s))]ds.$$

The inf is realized; let $\bar{v}(\cdot)$ be a measurable selection of the minimizer and define $\bar{V}(t) = x + \int_0^t \bar{v}(s)ds$. We have

$$S_G(T,\Phi) \geq \sum_{i=0}^{N-1} \sup_{\alpha} \int_{i\Delta}^{i\Delta+\Delta} [\alpha' v(s) - H(\alpha,\Phi(s))]ds. \tag{4.12}$$

The sups in (4.12) are attained at some $\{\alpha_i,\ 0 \leq i \leq N\}$. There are $s_i \in [i\Delta, i\Delta + \Delta)$ such that

$$\frac{1}{\Delta} \int_{i\Delta}^{i\Delta+\Delta} H(\alpha_i,\Phi(s))ds = H(\alpha_i,\Phi(s_i)).$$

Define $\psi_\Delta(\cdot)$ to be the function with value $\Phi(s_i)$ on $[i\Delta, i\Delta + \Delta)$. As $\Delta \to 0$, $\psi_\Delta(\cdot) \to \Phi(\cdot)$ uniformly on $[0,T]$. Thus

$$S_G(T,\Phi) \geq \sum_0^{N-1} \Delta \left[\frac{\alpha_i'[\bar{V}(i\Delta+\Delta) - \bar{V}(i\Delta)]}{\Delta} - H(\alpha_i,\psi_\Delta(i\Delta)) \right]$$

$$= \sum_0^{N-1} \Delta\, L\left[\frac{\bar{V}(i\Delta+\Delta) - \bar{V}(i\Delta)}{\Delta},\ \psi_\Delta(i\Delta) \right]. \tag{4.13}$$

Define $\Phi_\Delta(\cdot)$ to be the piecewise linear function with samples $\Phi_\Delta(0) = x$ and $\Phi_\Delta(i\Delta+\Delta) = \Pi_G(\Phi_\Delta(i\Delta) + \bar{V}(i\Delta+\Delta) - \bar{V}(i\Delta))$. Then by (4.13)

$$S_G(T,\Phi) \geq \sum_{i=0}^{N-1} \Delta \inf_{v\,:\,\Pi_G(\Phi_\Delta(i\Delta) + \Delta v) = \Phi_\Delta(i\Delta+\Delta)} L(v, \psi_\Delta(i\Delta))$$

$$= S_G^{\psi_\Delta,\Delta}(T,\Phi_\Delta)\ . \tag{4.11}$$

A proof similar to that of Theorem 3 below yields that $\Phi_\Delta(\cdot) \to \Phi(\cdot)$ uniformly on [0,T]. Thus (4.6) is proved. Q.E.D.

V. Large Deviations Results for $x^{\epsilon}(\cdot)$.

In order to extend Theorem 1 (for the $x^{\epsilon,\Delta}(\cdot)$ process) to the $x^{\epsilon}(\cdot)$ process, we need to show that the processes are close for small ϵ,Δ. Let $\Delta = k\epsilon$, k being a large integer. Recall the definitions

$$(5.1) \qquad X_{i+1}^{\epsilon,\Delta} = \Pi_G(X_i^{\epsilon,\Delta} + \epsilon \sum_{j\in I_i} b(X_i^{\epsilon,\Delta}, \xi_j)), \quad I_i = \{ik \le j < ik+k\},\ i \le T/\Delta$$

$$(5.2) \qquad X_{j+1}^{\epsilon} = \Pi_G(X_j^{\epsilon} + \epsilon b(X_j^{\epsilon},\xi_j)) \ , \ j \le T/\epsilon = kT/\Delta \ .$$

To extend the large deviations results to $x^{\epsilon}(\cdot)$ it is sufficient (Theorem 4 below) to show that for each $T < \infty$, $\phi(\cdot)$ and $\delta > 0$, there are $\delta_i > 0$ which tend to zero as $\delta \to 0$ such that for small enough Δ

$$(5.3) \qquad d(x^{\epsilon,\Delta}(\cdot),\phi(\cdot)) < \delta_1 \Rightarrow d(x^{\epsilon}(\cdot),\phi(\cdot)) < \delta$$

$$\Rightarrow d(x^{\epsilon,\Delta}(\cdot),\phi(\cdot)) < \delta_2$$

Let $\rho_j = b(\phi(j\epsilon),\xi_j)$ and define the processes $\{\bar{X}_n^{\epsilon}, n \le T/\epsilon\}$, $\{\bar{X}_{ik}^{\epsilon}, i \le T/\Delta\}$:

$$\bar{X}_{n+1}^{\epsilon} = \Pi_G(\bar{X}_n^{\epsilon} + \epsilon\rho_n) \ , \ \bar{X}_0^{\epsilon} = \bar{X}_0^{\epsilon,\Delta} = x,$$

$$\bar{X}^{\epsilon,\Delta}_{ik+k} = \Pi_G(\bar{X}^{\epsilon,\Delta}_{ik} + \epsilon \sum_{j\in I_i} \rho_j),$$

and their continous parameter interpolations $\bar{x}^{\epsilon}(\cdot)$ (interval ϵ) and $\bar{x}^{\epsilon,\Delta}(\cdot)$ (interval Δ).

By the convexity of G and the Lipschitz condition on $b(\cdot,\xi)$, given $\delta < 0$ (resp., $\delta' < 0$) there are δ_i (resp., δ_i') going to zero as δ (resp., δ') goes to zero and such that

$$d(\bar{x}^{\epsilon,\Delta}(\cdot),\Phi) < \delta_1 \Rightarrow d(x^{\epsilon,\Delta}(\cdot),\Phi) < \delta \Rightarrow d(\bar{x}^{\epsilon,\Delta}(\cdot),\Phi) < \delta_2$$

$$d(\bar{x}^{\epsilon}(\cdot),\Phi) < \delta_1' \Rightarrow d(x^{\epsilon}(\cdot),\Phi) < \delta' \Rightarrow d(\bar{x}^{\epsilon}(\cdot),\Phi) < \delta_2'$$

Thus, to show (5.3), we need only show that $d(\bar{x}^{\epsilon,\Delta}(\cdot), \bar{x}^{\epsilon}(\cdot)) \to 0$ as $\Delta \to 0$, $\Delta/\epsilon \to \infty$. We will actually bound $|\bar{X}^{\epsilon,\Delta}_{ik} - \bar{X}^{\epsilon}_{ik}|$, $\epsilon ik \leq T$.

For notational convenience, let $|\rho_j| \leq 1$ and absorb any other bound into the ϵ.

The basic idea is to show that if the two processes ever separate by $\Delta^{1/2}$, then the maximum rate of growth of the separation is then slow enough for them to stay close. The following lemma will be used the proof.

Lemma 2. Let x_1,x_2 be in G, with $v = x_1 - x_2$. Fix $\gamma > 0$. Let $y_1 \in N_\gamma(x_1) \cap G$, $y_2 \in N_\gamma(x_1)$, and $w = y_2 - y_1$. Then

$$\text{(5.4)} \qquad \langle \Pi_G(y_2) - y_2, \frac{v}{|v|} \rangle \leq \gamma |w| / |v| \ .$$

Proof. (See Fig. 2) If $y_2 \in G$, there is nothing to prove. Let $y_2 \notin G$. Consider the hyperplane defined by the normal $\Pi_G(y_2) - y_2$ and point y_2. Since G is convex, x_2 lies on the same side of this hyperplane as does $\Pi_G(y_2)$. Thus

$$\langle \Pi_G(y_2) - y_2, x_2 - y_2 \rangle \geq 0 \qquad \text{or}$$

$$\langle \Pi_G(y_2) - y_2, (x_2 - x_1) + (x_1 - y_2) \rangle \geq 0 \qquad \text{which implies}$$

$$\langle \Pi_G(y_2) - y_2, \frac{v}{|v|} \rangle \leq \frac{1}{|v|} \langle \Pi_G(y_2) - y_2, x_1 - y_2 \rangle \ .$$

Since $|\Pi_G(y_2) - y_2| \leq |w|$ and $|x_1 - y_2| \leq \gamma$, the lemma follows. Q.E.D.

<u>Theorem 3.</u> $\overline{\lim}_{\Delta}\ \overline{\lim}_{\epsilon}\ \sup_{ik \le T/\Delta} |\overline{X}^{\epsilon}_{ik} - \overline{X}^{\epsilon,\Delta}_{ik}| = 0.$

<u>Proof.</u> We use Lemma 2, where we identify $\overline{X}^{\epsilon}_{ik}$ with x_1 and $\overline{X}^{\epsilon,\Delta}_{ik}$ with x_2. Let $n \in [ik, ik + k)$ and set $y_1 = \overline{X}^{\epsilon}_{n}$, $y_2 = \overline{X}^{\epsilon}_{n} + \epsilon\rho_n$. Thus $\Pi_G(y_2) = \Pi_G(\overline{X}^{\epsilon}_{n} + \epsilon\rho_n)$. Since $|\rho_n| \le 1$ and $\Pi_G(\cdot)$ is a contraction and $k = \Delta/\epsilon$, we can use the value Δ for γ. Define $d_k = |\overline{X}^{\epsilon}_{ik} - \overline{X}^{\epsilon,\Delta}_{ik}|$. Then the lemma yields

$$<\Pi_G(\overline{X}^{\epsilon}_{n} + \epsilon\rho_n) - (\overline{X}^{\epsilon}_{n} + \epsilon\rho_n)\ ,\ \frac{\overline{X}^{\epsilon}_{ik} - \overline{X}^{\epsilon,\Delta}_{ik}}{d_k} > \le \epsilon\Delta/d_k$$

or, equivalently,

$$<\overline{X}^{\epsilon}_{n+1} - \overline{X}^{\epsilon}_{n} - \epsilon\rho_n\ ,\ \frac{\overline{X}^{\epsilon}_{ik} - \overline{X}^{\epsilon,\Delta}_{ik}}{d_k} > \le \epsilon\Delta/d_k\ .$$

Summing from ik to $ik + k - 1$ yields

$$<\overline{X}^{\epsilon}_{ik+k} - \overline{X}^{\epsilon}_{ik} - \epsilon\sum_{ik}^{ik+k-1}\rho_n\ ,\ \frac{\overline{X}^{\epsilon}_{ik} - \overline{X}^{\epsilon,\Delta}_{ik}}{d_k} > \le \Delta^2/d_k\ . \tag{5.5}$$

We next get an estimate for $\overline{X}^{\epsilon,\Delta}_{ik+k} - \overline{X}^{\epsilon,\Delta}_{ik}$. Use Lemma 2 again with $x_1 = \overline{X}^{\epsilon,\Delta}_{ik}$, $x_2 = \overline{X}^{\epsilon}_{ik}$, $y_1 = \overline{X}^{\epsilon,\Delta}_{ik}$, $y_2 = \overline{X}^{\epsilon,\Delta}_{ik} + \epsilon\sum_{ik}^{ik+k-1}\rho_n$. Then $\Pi_G(y_2) = \overline{X}^{\epsilon,\Delta}_{ik+k}$ and (using $\gamma = \Delta$)

$$<\overline{X}^{\epsilon,\Delta}_{ik+k} - \overline{X}^{\epsilon,\Delta}_{ik} - \epsilon\sum_{n=ik}^{ik+k-1}\rho_n\ ,\ \frac{\overline{X}^{\epsilon,\Delta}_{ik} - \overline{X}^{\epsilon}_{ik}}{d_k} > \le \Delta^2/d_k. \tag{5.6}$$

Subtracting (5.6) from (5.5) and defining $\bar{Y}_{ik}^{\epsilon,\Delta} = \bar{X}_{ik}^{\epsilon} - \bar{X}_{ik}^{\epsilon,\Delta}$ yields

(5.7) $$<\bar{Y}_{ik+k}^{\epsilon,\Delta} - \bar{Y}_{ik}^{\epsilon,\Delta}, \bar{Y}_{ik}^{\epsilon,\Delta}/d_k> \leq 2\Delta^2/d_k.$$

Suppose that $d_k \geq \Delta^{1/2}$. Then (5.7) implies that the component of $\bar{Y}_{ik+k}^{\epsilon,\Delta}$ in the direction $\bar{Y}_{ik}^{\epsilon,\Delta}$ has magnitude less than or equal to $2\Delta^2/d_k + d_k \leq 2\Delta^{3/2} + d_k$. The bound $|\rho_n| \leq 1$ implies that the projection of $\bar{Y}_{ik+k}^{\epsilon,\Delta}$ onto the hyperplane normal to $\bar{Y}_{ik}^{\epsilon,\Delta}$ has magnitude $\leq 2\Delta$. (In fact $|\bar{Y}_{ik+k}^{\epsilon,\Delta} - \bar{Y}_{ik}^{\epsilon,\Delta}| \leq 2\Delta$.) Thus, if $d_k \geq \Delta^{1/2}$,

$$d_{k+1}^2 \leq (2\Delta^{3/2} + d_k)^2 + 4\Delta^2.$$

Let k_1 denote the maximum distance across G. Then (for $d_k \geq \Delta^{1/2}$ and $2\Delta^{3/2} \leq 1$)

$$d_{k+1}^2 \leq (2 + 4\ k_1)\Delta^{3/2} + 4\Delta^2 + d_k^2$$

or, in general

$$d_k^2 \leq \max\left[\Delta^{1/2}, (6 + 4k_1)\frac{T}{\Delta} \cdot \Delta^{3/2}\right].$$

Q.E.D.

It follows from Theorems 2 and 3 that

Theorem 4. $S_G(T,\Phi)$ is an action functional for $\{x^\epsilon(\cdot)\}$ and (2.5) holds.

Proof. Fix the set A, and let $\Phi \in A^0$. Using (5.3) select $\delta > 0$, $\delta_1 > 0$ such that $N_\delta(\Phi) \subset A^0$ and for small Δ

$$\left\{d(x^{\epsilon,\Delta}(\cdot),\Phi(\cdot)) < \delta_1\right\} \subset \left\{d(x^\epsilon(\cdot),\Phi(\cdot)) < \delta\right\}.$$

It then follows that

$$\varliminf_{\epsilon} \epsilon \log P_x\left\{x^\epsilon(\cdot) \in A\right\}$$

$$\geq \varliminf_{\epsilon} \epsilon \log P_x\left\{d(x^\epsilon(\cdot),\Phi(\cdot)) < \delta\right\}$$

$$\geq \varliminf_{\Delta} \varliminf_{\epsilon} \epsilon \log P_x\left\{d(x^{\epsilon,\Delta}(\cdot),\Phi(\cdot)) < \delta_1\right\}$$

$$\geq - S_G^{\Delta}(T,\Phi)$$

where the last inequality is due to (3.8). This gives the left hand side of (2.5).

Since the estimates in Theorem 3 are independent of the particular Φ chosen we have

$$P_x\left\{x^\epsilon(\cdot) \in A\right\} \leq P_x\left\{x^{\epsilon,\Delta}(\cdot) \in N_\delta(A)\right\}$$

for any $\delta > 0$ and small enough ϵ, Δ. Hence by (3.8)

$$\overline{\lim_{\epsilon}} \ \epsilon \log P_x\{x^\epsilon(\cdot) \in A\}$$

$$\leq - \underline{\lim_{\delta}} \inf_{\phi \in N_\delta(A)} S_G(T,\phi).$$

Since by l.s.c.

$$\underline{\lim_{\delta}} \inf_{\phi \in N_\delta(A)} S_G(T,\phi)$$

$$\geq \inf_{\phi \in \bar{A}} S_G(T,\phi) ,$$

the right hand side of (2.5) is proved. Q.E.D.

VI. Mean Escape Time.

Let θ be an asymptotically stable point of (6.1), and D a neighborhood (relative to G) of θ with $\bar{D}$ in the domain of attraction of θ.

$$(6.1) \qquad \dot{x} = \Pi_G(x,\bar{b}(x))$$

Let τ_D^ϵ denote the escape time of $x^\epsilon(\cdot)$ from D. Then, under some additional assumptions, we will prove the analog of the classical case [6], [10], namely

$$(6.2) \qquad \lim_{\epsilon} \epsilon \log E_x \tau_D^\epsilon = S_D(\theta), \; x \in D,$$

where

$$S_D(\theta) = \inf \{S_G(T,\phi): \phi(0) = \theta, \phi(T) \in \partial D\} \quad .$$

All neighborhoods are relative to G.

In order to avoid excess detail, we work with the non-degenerate case, (see below (2.6)) but the results hold for the non-degenerate case as well, if we assume the existence of the $\phi^x(\cdot)$ discussed below (6.3).

Since

$$L_G(\beta,x) = \inf_{u \,:\, \Pi_G(x,\bar{b}(x) + u) = \beta} L(\bar{b}(x) + u,x)$$

and $L(\alpha,x) = 0$ if and only if $\alpha = \bar{b}(x)$, we have

$$L_G(\beta,x) = 0 \quad \text{iff} \quad \beta = \Pi_G(x,\bar{b}(x)) \quad . \tag{6.3}$$

Loosely speaking, we 'pay' only when noise or 'control' u is required to force a deviation from the (free) path of (6.1).

For each small $\rho_1 > 0$ and $N_\rho(\theta)$ there is a ρ_2 ($\rho_2 \to 0$ as $\rho_1 \to 0$) and a T_0 such that all paths of (6.1) starting in $\bar{D}$ reach $N\rho_1(\theta)$ by time T_0 and do not leave $N\rho_2(\theta)$ after first hitting $\partial N\rho_1(\theta)$. By the non-degeneracy assumption $L(\bar{b}(x) + u,x)$ is strictly convex in u and equals $o(u)$ uniformly in $x \in G$. This implies the following. For each $\rho > 0$ there are $T < \infty$ and $\rho_1 > 0$ ($\rho_1 \to 0$ as $\rho \to 0$) such that for each $x \in N\rho_1(\theta)$ there is a path $\phi^x(\cdot)$ such that $\phi^x(0) = x$, $\phi^x(t^x) = \theta$ for some $t^x \leq T$ and $S_G(t^x, \phi^x) \leq \rho$.

It is sometimes convenient to define ϕ^x for $t > t^x$ without increasing the cost. To do this, we let $\phi^x(\cdot)$ satisfy (6.1) beyond t^x. For suitable ρ_2 (going to zero as ρ and $\rho_1 \to 0$), we can suppose that $\phi^x(\cdot)$ never leaves $N\rho_2(\theta)$.

Define (if the set is empty, define the inf to be ∞)

$$S_D(x) = \inf \{S_G(T,\phi) : \phi(0) = x,\ \phi(T) \notin D,\ T < \infty\}.$$

Given $\delta > 0$, there is a $T^\delta < \infty$ and for each $x \in D$ a path $\tilde{\phi}^x(\cdot)$ on the interval $[0,T^\delta]$ such that $\tilde{\phi}^x(0) = x$, $\tilde{\phi}^x(t^x) \in \partial D$ at some $t^x \leq T^\delta$, $\tilde{\phi}^x(\cdot)$ satisfies (6.1) after t^x and $S_G(T^\delta,\tilde{\phi}^x) \leq S_D(x) + \delta$. This fact, which will be used frequently in the sequel, follows from the following two observations:

(1) For each $\rho > 0$ and any set of paths in G with uniformly bounded costs, there is a $T_\rho < \infty$ such that the paths must spend all but at most T_ρ units of time in $N_\rho(\theta)$;

(2) There is a $T'_\rho < \infty$ and paths $\phi_1^x(\cdot)$ on the interval $[0,T'_\rho]$ taking $x \in N_\rho(\theta)$ and then to θ and then to ∂D at some time $\leq T'_\rho$, with cost $\leq S_D(\theta) + \rho_1$, where $\rho_1 \to 0$ as $\rho \to 0$.

We will next show that $S_D(x) \to S_D(\theta)$ as $x \to \theta$. By the comments in the previous paragraphs, $S_D(x) \leq S_D(\theta)$ for $x \in D$, since x can be connected to θ by a path with arbitrarily small cost. If $\lim_{y \to \theta} S_D(y) = \infty$, then we are done. Thus, we need only work with sequences $x_n \to \theta$ such that $\sup_n S_D(x_n) < \infty$. Let $x_n \to \theta$ and fix $\delta > 0$. There are $\phi_n^\delta(\cdot)$ and bounded T_n^δ (by, say $\bar{T}$) such that $\phi_n^\delta(0) = x_n$, $\phi_n^\delta(T_n^\delta) \in \partial D$ and $S_G(T_n^\delta,\phi_n^\delta) \leq S_D(x_n) + \delta$. For $\bar{T} \geq t > T_n^\delta$, let $\phi_n^\delta(\cdot)$ satisfy (6.1). Choose (and index by n) a convergent subsequence of $\{T_n^\delta, \phi_n^\delta(\cdot)\}$ with limit $\{T,\phi^\delta(\cdot)\}$. Then $\phi^\delta(0) = \theta$ and $\phi^\delta(T) \in \partial D$. By the l.s.c. of $S_G(\bar{T},\cdot)$,

$$\delta + \underline{\lim}_n\, S_D(x_n) \geq \underline{\lim}_n\, S_G(\bar{T},\phi_n^\delta) \geq S_G(\bar{T},\phi) = S_G(T,\phi) \geq S_D(\theta).$$

Thus, since δ is arbitrary

$$\lim_{x \to \theta} S_D(x) = S_D(\theta). \tag{6.3}$$

<u>Assumptions</u>. We carry over the assumptions from Sections 1 and 2. For the escape time problem one must <u>redefine</u> the $\bar{b}(x)$ and $H(\alpha,x)$ of (2.1) and (2.2). Let B_m be the minimal σ-algebra measuring $\{\xi_i, i \le m\}$ and if M is a stopping time for $\{\xi_m\}$, let B_M denote the associated σ-algebra. If τ is a stopping time for the continuous parameter process $x^\epsilon(\cdot)$, we use $B_\epsilon(\tau)$ instead of $B_{[\tau/\epsilon]}$. Suppose that the limits in (6.4) and (6.5) exist uniformly in the stopping time M, in ω, and in (x,α) in any compact set, and that $H(\cdot,x)$ is differentiable. <u>Those properties hold for the processes listed below</u> (2.2).

(6.4) $$\bar{b}(x) = \lim_n \frac{1}{n} E_{B_M} \sum_{M}^{n+M-1} b(x,\xi_j)$$

(6.5) $$H(\alpha,x) = \lim_n \frac{1}{n} \log E_{B_M} \exp \alpha' \sum_{M}^{n+M-1} b(x,\xi_j) ,$$

For τ a stopping time for $x^\epsilon(\cdot)$, let $P_{x,B_\epsilon(\tau)}$ denote the conditional probability measure of the process $x^\epsilon(\cdot)$ <u>which is reset to</u> x <u>at time</u> τ, then evolves as before (using $\{\xi_{\tau/\epsilon+j}, j \ge 0\}$ after time τ), <u>conditioned on the data</u> $B_\epsilon(\tau)$ <u>up to</u> τ. Define $S_G(T,A) = \inf_{\phi\in A} S_G(T,\phi)$. Then for each $T < \infty$ and $A \subset C_x[0,T]$, we have

(6.6a) $$-S_G(T,A^0) \le \underline{\lim}_{\epsilon} \epsilon \log P_{x,B_\epsilon(\tau)}(x^\epsilon(\tau+\cdot) \in A)$$

$$\le \overline{\lim}_{\epsilon} \epsilon \log P_{x,B_\epsilon(\tau)}(x^\epsilon(\tau+\cdot) \in A) \le -S_G(T,\bar{A})$$

The 'rate' at which the inequalities hold is uniform in τ and ω in the sense that (e.g), for each $h > 0$ there is $\epsilon_0 > 0$ such that for all τ,ω and $\epsilon \leq \epsilon_0$,

$$(6.6b)\ \exp - [S_G(T,A^0) + h]/\epsilon \leq P_{x,B_\epsilon(\tau)}\{x^\epsilon(\tau + \cdot) \in A\} \leq \exp - [S_G(T,\bar{A}) - h]/\epsilon.$$

The 'uniformity' in (6.6) follows from the uniformity of the convergence in (6.4) and (6.5) in the same variables (ω,M). In fact, our derivation started with (3.2), (3.3), obtainable from [6]. It follows from the derivation in [6] (although not mentioned explicitly there) that the probability in (3.3) can be replaced by $P_{x,B_M}\{\{Y_i^{\epsilon,\psi,\Delta}, i \leq N\} \in B\}$ with the 'rate' at which the inequalities hold being uniform in all variables (ω,M) in which the convergence in (6.4), (6.5) is uniform. Here, P_{x,B_M} denotes the probability measure (conditioned on B_M) of $\{X_n^\epsilon\}$ (or $\{Y_i^{\epsilon,\psi,\Delta}\}$) <u>reset</u> to x at time M, then evolving as before (using $\xi_{j+M}, j \geq 0$, after M).

We will make one additional assumption. Let D_δ denote a δ-neighborhood of D with $D_0 = D$. Then, clearly, $S_{D_\delta}(\theta)$ decreases as $\delta \downarrow 0$. We assume that $S_{D_\delta}(\theta) \downarrow S_D(\theta)$ as $\delta \downarrow 0$. If this condition doesn't hold for D it will hold for an arbitrarily small perturbation of D. If θ lies in the interior of G and if the optimal exit path does not hit ∂G, then this condition is implied by the non-degeneracy assumption.

<u>Theorem 5.</u> <u>Under the assumptions in the above subsection</u>, (6.2) <u>holds.</u>

<u>Proof.</u> <u>Part 1.</u> We follow [10] as closely as possible and omit details when they are sufficiently close to those in [10]. Assume $S_D(\theta) < \infty$. Otherwise, a similar proof yields the result (in fact, $x^\epsilon(\cdot)$ cannot then escape D with full probability for small ϵ). Let $0 < \mu_1 < \mu_2 < \mu_3$. Define $g_0 = N_{\mu_1}(\theta)$, $\Gamma_0 = N_{\mu_3}(\theta) - N_{\mu_2}(\theta)$, with all $N_{\mu_i}(\theta)$ contained in D. Define the stopping times $\{\sigma_i, \tau_i\}$ by $\tau_0 = 0$ and

$$\sigma_n = \inf \{ t > \tau_n : x^\epsilon(t) \in \Gamma_0 \}$$

$$\tau_n = \inf \{ t > \sigma_{n-1} : x^\epsilon(t) \in g_0 \cup (G-D) \},$$

and set $Z_n = x^\epsilon(\tau_n \cap \Gamma_0^\epsilon)$. For rotational simplicity, we omit the ϵ-dependence on σ_n, τ_n, Z_n. We have (for $x \in g_0$, otherwise we add a term $E_x \tau_1$, which is bounded uniformly in x and ϵ, to (6.7))

$$\text{(6.7)} \qquad E_x \tau_D^\epsilon = \sum_0^\infty E_x I_{\{Z_n \in g_0\}} E_{Z_n, B_\epsilon(\tau_n)} (\tau_{n+1} - \tau_n)$$

The theorem will be proved via estimates of the terms in (6.7).

We have, for $x \in g_0$,

$$\text{(6.8)} \qquad \inf_{y \in \Gamma_0, \omega, n} E_{y, B_\epsilon(\sigma_n)} (\tau_{n+1} - \sigma_n) \leq E_{x, B_\epsilon(\tau_n)} (\tau_{n+1} - \tau_n)$$

$$\leq \sup_{y \in \Gamma_0, \omega, n} E_{y, B_\epsilon(\sigma_n)} (\tau_{n+1} - \sigma_n) + \sup_{y \in g_0, \omega, n} E_{y, B_\epsilon(\tau_n)} (\sigma_n - \tau_n).$$

It can be shown that there are $k_i > 0$ (depending on the μ_i) such that the left side of (6.8) is bounded below by k_1 and the first term on the r.h.s. is bounded above by k_2 (the latter fact follows from an argument similar to that which uses Lemma 3 in Part 3 below).

Let $d > 0$. Let $\hat{\phi}(\cdot)$ denote a d/4-optimal path from θ to $\partial N_{\mu_3}(\theta)$, and write $h = \mu_3 - \mu_2$. For small μ_i, there is a $T < \infty$, (depending on μ_i but not on x) such that for each $x \in N_{\mu_2}(\theta)$, there is a path $\hat{\phi}^x(\cdot)$ taking x to θ with (cost $\leq d/4$) then (using the first part of $\hat{\phi}(\cdot)$ here) θ to $\partial N_{\mu_3}(\theta)$, at a total cost no greater than $d/2$. Then, there is an $\tilde{\epsilon} > 0$ such that for $\epsilon \leq \tilde{\epsilon}$ and all ω, n

$$\text{(6.9)} \qquad P_{x,B_\epsilon(\tau_n)}\{d(x^\epsilon(\tau_n+\cdot), \hat{\phi}^x(\cdot)) < h/2\} \geq \exp - d/\epsilon .$$

The $\tilde{\epsilon}$ can be chosen to be independent of $x \in N_{\mu_2}(\theta)$, although we omit the details, (The argument is similar to that used below to get the uniform bound on the terms in the sum in (6.16).)

Let $\rho + \tau_n = \sigma_n$ denote the first escape time into Γ_0 after τ_n. Then (in this calculation, we let $x^\epsilon(\tau_n) \in g_0$, but for simplicity we omit the associated notation)

$$\text{(6.10)} \qquad E_{x,B_\epsilon(\tau_n)}\rho \leq T \sum_{m=1}^{\infty} P_x\{\rho > mT\} .$$

By (6.9)

$$P_{x,B_\epsilon(\tau_n)}\{\rho > mT + T\} = E_{x,B_\epsilon(\tau_n)}\Big[1-P_{x^\epsilon(\tau_n+mT),\,B_\epsilon(\tau_n+mT)}\{\rho - mT \le T\}\Big]I_{\{\rho > mT\}}\ .$$

$$\le E_{x,B_\epsilon(\tau_n)}[1 - \exp - d/\epsilon]\ I_{\{\rho > mT\}}$$

$$\le [1 - \exp - d/\epsilon]^{m+1}.$$

Thus $E_{x,B_\epsilon(\tau_n)}\rho \le T \exp d/\epsilon$ for small ϵ.

Putting these estimates together yields that (up to a multiplicative factor in $[k_1, k_2 + T\exp d/2]$ for arbitrarily small d), (6.7) equals

$$\sum_0^\infty P_x\{Z_n \in g_0\}\ , \tag{6.11}$$

which we evaluate next.

Part 2. Fix $d > 0$. For small μ, there are $\delta_0 > 0$, $t_1 < \infty$ and $h > 0$ such that for each $x \in g_0$ there is a function $\bar\phi^x(\cdot)$ on $[0,t_1]$ connecting x to Θ, then Θ to $\partial D_h = \partial N_h(D)$ at some time $t^x \le t_1$ with the following properties: $\bar\phi^x(\cdot)$ satisfies (6.1) after t^x; $S_G(t^x,\bar\phi^x) = S_G(t_1,\bar\phi^x) \le S_D(\Theta) + d/4$; the distance from the set g_0 of the part of the path from first exit of Γ_0 to first reaching ∂D_h is $\ge \delta_0$; the distance from the set Γ_0 of the part of the path which connects x to Θ is $\ge \delta_0$. A similar construction was used in [10, p 124]. The fact that the minimum cost for hitting ∂D_h is close to that for hitting ∂D (for small h) follows from the last assumption stated above Theorem 5. Let $\delta_1 = \min(\delta_0, h)$. Then there is an $\epsilon_0 > 0$ such that for $\epsilon \le \epsilon_0$ (we compare functions on the interval $[0,t_1]$ here) and $x \in g_0$,

$$P_{x,B_\epsilon(\tau_n)}\{Z_{n+1} \notin D\} \geq \tag{6.12}$$

$$P_{x,B_\epsilon(\tau_n)}\{d(x^\epsilon(\tau_n + \cdot), \bar{\Phi}^x(\cdot)) \le \delta_1\}$$

$$\geq \exp - [S_D(\Theta) + d/2] \;/\epsilon \;.$$

As for (6.9), ϵ_0 can be chosen independently of $x \in g_0$. We have

$$P_x\{Z_{n+1} \in g_0\} = P_x\{\tau_D^\epsilon > \tau_{n+1}\} = E_x[1 - P_{x^\epsilon(\tau_n),B_\epsilon(\tau_n)}\{Z_{n+1} \in g_0\}] I_{\{Z_n \in g_0\}}$$

Using (6.12) to get an upper bound on the bracketed terms and iterating yields

$$P_x\{Z_{n+1} \in g_0\} \le [1 - \exp - (S_D(\Theta) + d/2)\;/\epsilon]^{n+1} \tag{6.13}$$

which yields the upper bound $\exp [S_D(\Theta) + d/2]\;/\epsilon$ on the sum in (6.7) when the $(\tau_{n+1} - \tau_n)$ terms are dropped from the sum).

<u>Part 3.</u> To complete the proof, we need the following Lemma, whose proof is very similar to that of Lemma 1.9 of [10, Chapter 6] and is omitted.

<u>Lemma 3.</u> <u>Let</u> K <u>be a compact set in</u> G <u>which does not contain an entire limit set for</u> (6.1), <u>and let</u> τ <u>denote a stopping time for</u> $x^{\epsilon}(\cdot)$. <u>Define</u> $\tau_K^{\epsilon} = \min\{t: x^{\epsilon}(\tau + t) \notin K\}$. <u>Then there are</u> $c > 0$, $T_0 < \infty$, $\epsilon_0 > 0$ <u>such that for</u> $\epsilon \le \epsilon_0$ <u>and all</u> $y \in K$ <u>and all</u> T

$$P_{y,B_\epsilon(\tau)}\{\tau_K^{\epsilon} > T\}$$

$$\le \exp - c\,(T - T_0)/\epsilon$$

<u>for all</u> τ,ω.

Continuing with the proof of the Theorem, we have for any $t_2 < \infty$

$$(6.14) \qquad P_{x,B_\epsilon(\tau_n)}\{Z_{n+1} \notin g_0\} = P_{x,B_\epsilon(\tau_n)}\{Z_{n+1} \notin D\} \le$$

$$\sup_{y \in \Gamma_{0,\omega}} P_{y,B_\epsilon(\sigma_n)}\{Z_{n+1} \notin D\}$$

$$\le \sup_{y \in \Gamma_{0,\omega}} P_{y,B_\epsilon(\sigma_n)}\{\tau_{n+1} - \sigma_n \ge t_2\}$$

$$+ \sup_{y \in \Gamma_{0,\omega}} P_{y,B_\epsilon(\sigma_n)}\{\tau_{n+1} - \sigma_n \le t_2,\ Z_{n+1} \notin D\}$$

By Lemma 3, for any $k_3 < \infty$, there is a $t_2 < \infty$ such that for small ϵ the first term after the inequality of (6.14) is $\le \exp - k_3/\epsilon$. Fix $d > 0$. We next show that

$$(6.15)\qquad \sup_{x \in \Gamma_0, \omega} P_{x,B_\epsilon(\sigma_n)}\{\tau_{n+1} - \sigma_n \leq t_2, Z_{n+1} \notin D\} \leq \exp - (S_D(\Theta) - d)/\epsilon$$

for small $\epsilon > 0$.

Let $\sup_i |\xi_i| \leq k_4$. The set Q of all piecewise linear interpolations of all paths $x^\epsilon(\cdot)$ on $[0,t_2]$ (over all ϵ, initial conditions $x \in \bar{\Gamma}_0$, and an arbitrary k_4-bounded sequence $\{\xi_i\}$ used in lieu of $\{\xi_i\}$) is equicontinous. Let Q_x denote the closure of the subset of functions in Q with initial condition x, and which hit ∂D at some $t \leq t_2$. For small μ_1 and $\phi(\cdot) \in Q_x$, $S_G(t_2,\phi) \geq S_D(\Theta) - d/4$. Given $\delta > 0$, there are $N_\delta < \infty$ (not *depending on* $x \in \bar{\Gamma}_0$) and $\{\phi_i^x(\cdot), i \leq N_\delta\}$ in Q_x forming a $\delta/4$-net on Q_x. Note that if $x_n \to x$ and $\phi_{i_n}^{x_n}(\cdot) \to \phi(\cdot)$, then $\phi(\cdot) \in Q_x$. Now,

$$(6.16)\qquad \sup_\omega P_{x,B_\epsilon(\sigma_n)}\{\tau_{n+1} - \sigma_n \leq t_2, Z_{n+1} \notin D\}$$

$$\leq \sum_{i=1}^{N_\delta} \sup_\omega P_{x,B_\epsilon(\sigma_n)}\{d(x^\epsilon(\cdot),\phi_i^x(\cdot)) \leq \delta/2\}$$

We now show that the r.h.s. of (6.16) can be bounded independently of $x \in \bar{\Gamma}_0$. By (6.6) for each $x \in \bar{\Gamma}_0$ there is an $\tilde{\epsilon}(x) > 0$ such that for $\epsilon \leq \tilde{\epsilon}(x)$

$$\sup_{\omega, i} P_{x,B_\epsilon(\sigma_n)}\{d(x^\epsilon(\cdot), \phi_i^x(\cdot)) \leq \delta/2\} < \exp - (S_D(\Theta) - d)/\epsilon .$$

If $\inf_{x \in \bar{\Gamma}_0} \tilde{\epsilon}(x) = 0$, then there are $x \in \bar{\Gamma}_0$, $\phi(\cdot)$, $\epsilon_m \to 0$, $x_m \to x$, $\{i_m\}$, $\phi_{i_m}^{x_n}(\cdot) \to$

$\Phi(\cdot)$ such that on a set of positive probability for each m,

$$(6.17) \qquad P_{x_m, B_{\epsilon_m}(\sigma_n)}\{d(x^{\epsilon_m}(\cdot), \phi^{x_m}_{i_m}(\cdot)) \le \delta/2\} \ge \exp - (S_D(\Theta) - d)/\epsilon_m .$$

Again, by (6.6), there is an $\hat{\epsilon} > 0$ such that for $\epsilon \le \hat{\epsilon}$ and large m and all ω

$$(6.18) \qquad \exp - (S_D(\Theta) - d/2)/\epsilon \ge P_{x,B_\epsilon(\sigma_n)}\{d(x^\epsilon(\cdot),\Phi(\cdot)) \le \delta\}$$

$$\ge P_{x_m, B_\epsilon(\sigma_n)}\{d(x^\epsilon(\cdot),\phi^{x_m}_{i_m}(\cdot)) \le \delta/2\} .$$

This contradicts (6.17). Thus, we can bound the r.h.s. of (6.16) above by

$$(6.19) \qquad N_\delta \exp - (S_D(\Theta) - d)/\epsilon$$

for all x,ω,n and small ϵ.

Define $\nu = \min\{n : Z_n \notin g_0\}$. For small ϵ,

$$P_x\{\nu > n+1\} = P_x\{Z_j \in g_0, \text{ all } j \le n+1\}$$

$$= E_x P_{x^\epsilon(\sigma_n),B_\epsilon(\sigma_n)}\{Z_{n+1} \in g_0\} I_{\{\nu > n\}}$$

(6.20)

$$\ge \Big[\inf_{y \in \Gamma_0,\omega,n} P_{y,B_\epsilon(\sigma_n)}\{Z_{n+1} \in g_0\}\Big] P_x\{\nu > n\}$$

$$\ge (1 - \exp - [S_D(\Theta) - 2d]/\epsilon)^{n+1}.$$

This, together with (6.7), (6.13) and the arbitrariness of d yields the theorem since the $E_{s_n, B_\epsilon}(\tau_n)(\tau_{n+1} - \tau_n)$ values lie in the interval $[k_1, k_2 + T\exp d_1/\epsilon]$ for arbitrarily small d_1, as shown above. Q.E.D.

7. Remarks and Extensions.

7.1 Exit points of $x^\epsilon(\cdot)$ from D. Let there be a finite number of points $y_1,\dots, y_q \in \partial D$ such that

$$\inf_{T>0} \inf_{\phi \in A_i} S_G(T,\phi) = \inf_{T>0} \inf_{\phi \in A} S_G(T,\phi),$$

where $A_i = \{\phi(\cdot) \in C_\Theta[0,T] : \phi(T) = y_i\}$, $A = \{\phi(\cdot) \in C_\Theta[0,T]: \phi(T) \in \partial D\}$. Then, as in [10], for each $x \in D$ and $\delta > 0$,

$$\lim_{\epsilon \to 0} P_x\{d(x^\epsilon(\tau_D^\epsilon), \bigcup_1^g y_i) < \delta\} \to 1.$$

7.2. Global behavior of $x^\epsilon(\cdot)$ on $[0,\infty]$. Assume the non-degenerate case. Let $K_1,\dots, K_m$ denote a collection of disjoint compact sets, each one of which is a limit set for (1.2), and such that $\bigcup_i^m K_i$ contains all the limit sets for (1.2). If $K_i \cap \partial G \neq \phi$, let $K_i = \Theta_i$, a single point. For a diffusion with small noise, [10, chapter 7] obtains the (asymptotic) probabilities of transition from a neighbohood of K_i to one of K_j, and the (asymptotic) mean times spent in a neighborhood of any subset of $\{K_i, i \leq m\}$, before exiting to a neighborhood of another subset of the $\{K_i, i \leq m\}$.

Although our process is not Markov, similar results can be obtained here. Let g_j denote a μ_1-neighborhood of K_j and let Γ_j denote the set $N_{\mu_3}(K_j) - N_{\mu_2}(K_j)$. Define $\tau_0 = 0$, $\sigma_n = \inf\{t > \tau_n: x^\epsilon(t) \in \bigcup_i \Gamma_i\}$ and $\tau_n = \inf\{t > \sigma_{n+1}: x^\epsilon(t) \in \bigcup_i g_i\}$. Set $Z_n = x^\epsilon(\tau_n)$. Via the methods in the last section, one can get upper and lower estimates for $P_{x,B_\epsilon(T)}^{n}(Z_{n+1} \in g_j)$ for $x \in g_i$. These would then be used to obtain the results of [10, Chapter 7] exactly as the $P_x(Z_{n+1} \in g_j)$ are used in the Markov process case of that reference. All the limit expressions carry over, with use of our action functional $S_G(T,\Phi)$ in lieu of the action functional $S_{OT}(\Phi)$ of [10].

7.3. Stochastic approximation. Let $a_n > 0$, $a_n \to 0$, $\sum a_n = \infty$. The results of Sections 1 to 5 can be carried over to the projected stochastic approximation (SA)

$$X_{j+1} = \Pi_G(X_j + a_j b(X_j,\xi_j)),$$

where we use the conditions on $\{\xi_j\}$ and $b(\cdot,\cdot)$ of Section 2. Define $t_n = \sum_0^{n-1} a_i$ and the shifted processes

$$X_{j+1}^n = \Pi_G(X_j^n + a_j b(X_j^n,\xi_j)),\ j \geq n,\ X_n^n = x,$$

$$x^n(t) = \frac{X_{j+1}^n(t - t_j + t_n) - X_j^n(t - t_{j+1} + t_n)}{t_{j+1} - t_j} \quad \text{on} \quad [t_j - t_n, t_{j+1} - t_n),$$

$$\tau_D^n = \min\{t: x^n(t) \notin D\}.$$

References [8], [13] deal with the (unprojected) SA problem via large deviations. It is easy to incorporate the method of [8] with the 'projected' case of this paper, by accounting for the 'time varying' scaling $\{a_n\}$. We cite only one result (the Kiefer-Wolfowitz case can also be treated).

For $a_n = 1/n$, use the action functional

$$S_G(T,\phi) = \int_0^T e^s L_G(\dot{\phi}_s,\phi_s)ds,$$

and for $a_n = 1/n^{\alpha}$, $\alpha < 1$, use $S_G(T,\Phi) = \int_0^T L_G(\Phi_s,\dot{\Phi}_s)ds$. Then for $A \subset C_x[0,T]$,

$$- \inf_{\Phi \in A^0} S_G(T,\Phi) \leq \underline{\lim}_n \, a_n \log P_x\{x^n(\cdot) \in A\}$$

$$\leq \overline{\lim}_n \, a_n \log P_x\{x^n(\cdot) \in A\}$$

$$\leq - \inf_{\Phi \in \bar{A}} S_G(T,\Phi) .$$

Let $A = \{\Phi(\cdot) : \Phi(0) = x, \Phi(t) \notin D, \text{ some } t \leq T\}$, $x \in D$, where we define Θ and D as Section 6. Then, under the 'continuity' condition on ∂D just above Theorem 5,

$$\lim_{x \to \Theta} \lim_n \, a_n \log P_x\{\tau_D^n \leq T\} = - \inf_{\Phi \in \bar{A}} S_G(T,\Phi).$$

References

[1] Kushner, H.J. and D.S. Clark, Stochastic Approximation Methods for Constrained and Unconstrained Systems, 1978, Springer Verlag, Berlin.

[2] Kushner, H.J. and A. Shwartz, "An invariant measure approach to the convergence of stochastic approximations with state-dependent noise," SIAM J. on Control and Optim., 22 (Jan. 1984), 13-27.

[3] Ermoliov, Yu, Methods of Stochastic Programming, Nauka, Moscow, 1976 (in Russian).

[4] Kushner, H.J. and A. Schwarz, "Stochastic Approximation in Hilbert Space: Identification and optimization of linear continuous parameter systems", to appear SIAM J. on Control and Optimization.

[5] Pflug, G., "Stochastic minimization with constant step-sizes", to appear in SIAM J. on Control and Optimization.

[6] Freidlin, M.I., "The Averaging Principle and Theorems on Large Deviations", Russian Math. Surveys, 33 (July-Dec., 1978), 117-176.

[7] Kushner, H.J., "Robustness and approximation of escape times and large deviations estimates for systems with small noise effects', SIAM J. Appl. Math, 44, 1984, p.160-182.

[8] Dupuis, P. and H.J. Kushner, "Stochastic Approximation via Large Deviations: Asymptotic Properties", to appear in SIAM J. on Control and Optimization, September 1985.

[9] Varadhan, S.R.S., Large Deviations and Applications, CBMS-NSF Regional Conference Series, SIAM Philadelphia, 1984.

[10] Freidlin, M.I. and A.D. Ventsell, Random Perturbations of Dynamical Systems, Springer, Berlin, 1984.

[11] Parthasarathy, T., Selection Theorems and Their Applications, Lecture Notes in Math. 263, Springer-Verlag, 1972.

[12] Kushner, H.J., Approximation and Weak Convergence Methods for Random Processes: with Applications to Stochastic Systems Theory, MIT Press, Cambridge, Mass. USA, 1984.

[13] Korostelev, A.P., Stochastic Recurrent Processes, Nauka, Moscow, 1984.

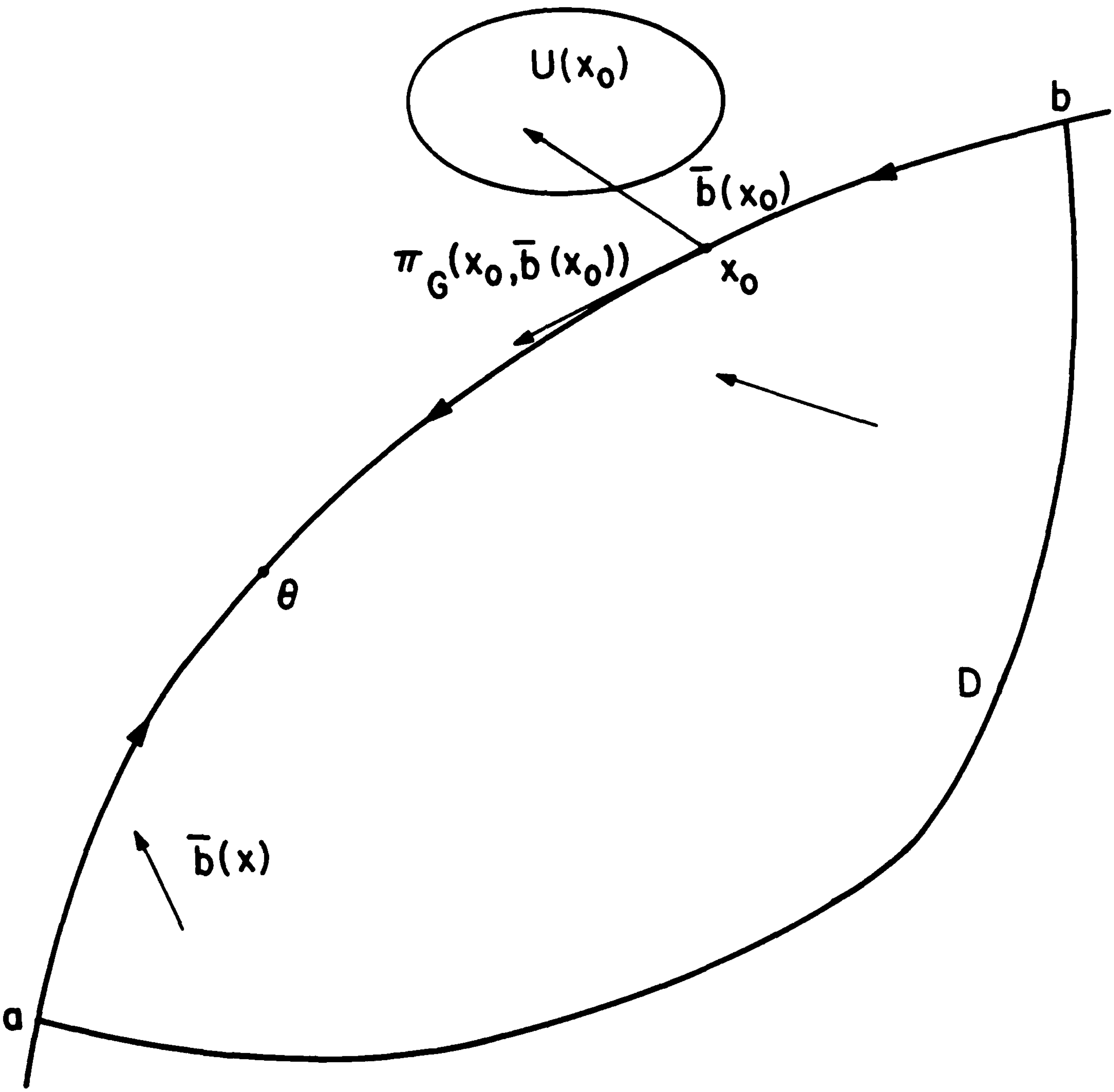


Fig. 1 Example of Escape and Flow Lines.
P(Escape is along boundary) $\xrightarrow{\epsilon}$ 1.

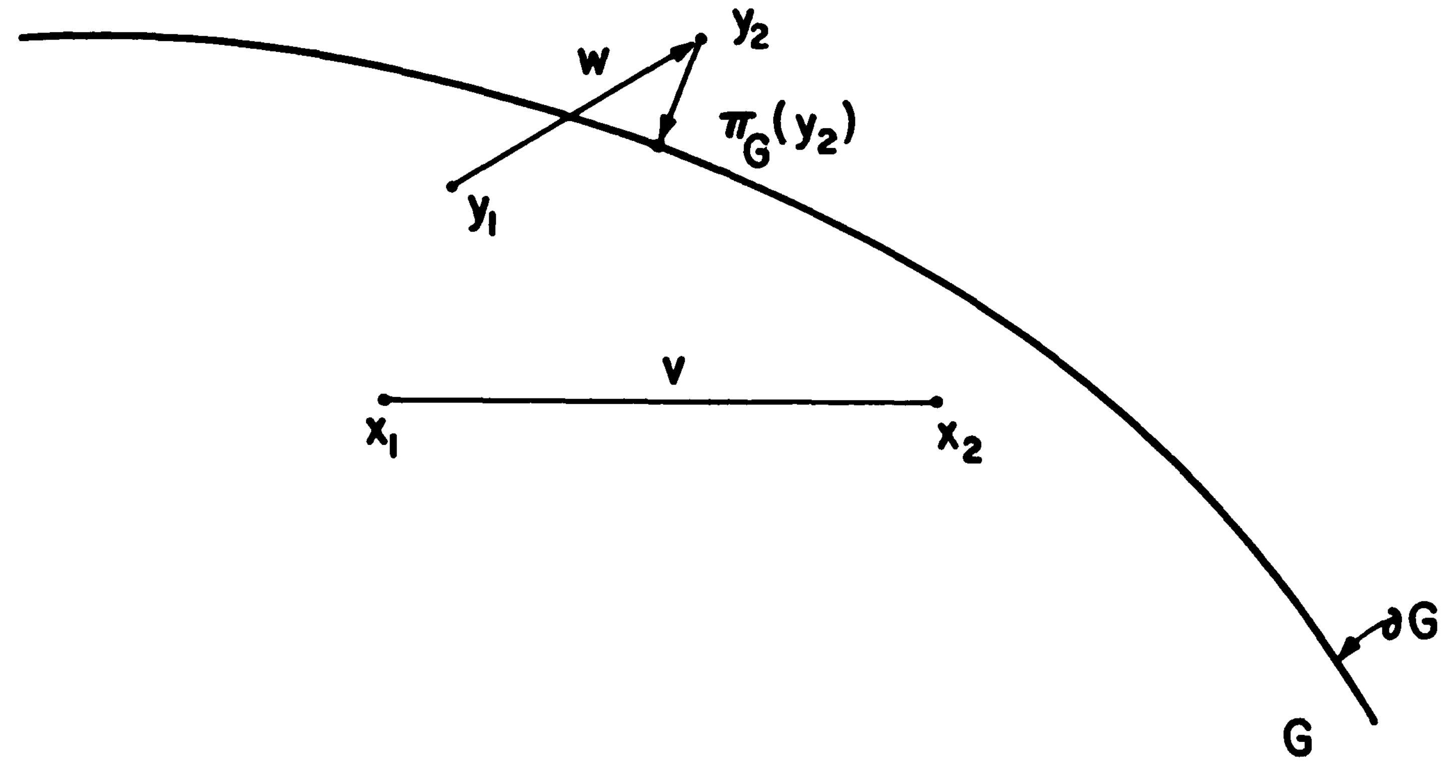


Fig.2 Illustration for Lemma 2.

END

FILMED

1-86

DTIC